Comparison of Latin Hypercube and Quasi Monte Carlo Sampling Techniques

Sergei Kucherenko[1], Daniel Albrecht[2], Andrea Saltelli[2]

[1]Imperial College London, UK
Email: s.kucherenko@imperial.ac.uk
[2]The European Commission, Joint Research Centre, ISPRA(VA), ITALY

Outline

Monte Carlo integration methods

Latin Hypercube sampling design

Quasi Monte Carlo methods. Sobol' sequences and their properties

Comparison of sample distributions generated by different techniques

Do QMC methods loose their efficiency in higher dimensions ?

Global Sensitivity Analysis and Effective dimensions

Comparison results

Monte Carlo integration methods

$$I[f] = \int_{H^n} f(\vec{x})d\vec{x}$$

see as an expectation: $I[f] = E[f(\vec{x})]$

Monte Carlo : $I_N[f] = \frac{1}{N}\sum_{i=1}^{N} f(\vec{z}_i)$

$\{\vec{z}_i\}$ – is a sequence of random points in H^n

Error: $\varepsilon = \left| I[f] - I_N[f] \right|$

$$\varepsilon_N = (\,E(\varepsilon^2))^{1/2} = \frac{\sigma\ (f)}{N^{1/2}} \rightarrow$$

Convergence does not depent on dimensionality but it is slow

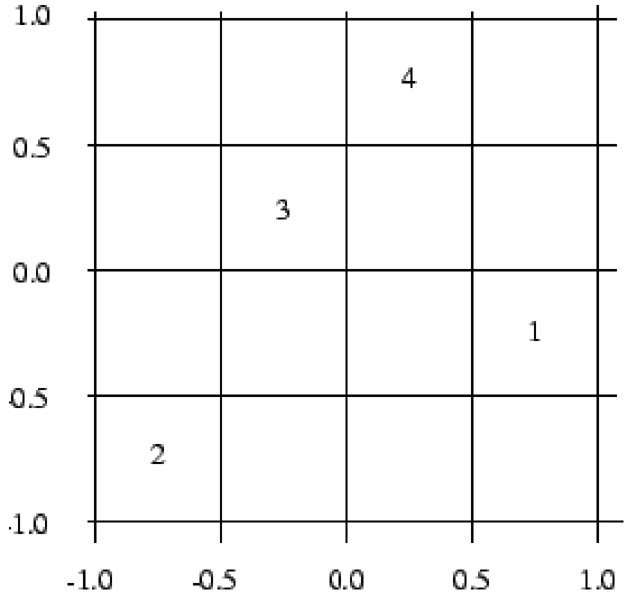
Improve MC convergence by decreasing $\sigma\ (f)$

Use variance reduction techniques:

antithetic variables; control variates;

stratified sampling $\rightarrow$ LHS sampling

Latin Hypercube sampling



Latin Hypercube sampling is a type of Stratified Sampling.

To sample N points in d-dimensions

Divide each dimension in N equal intervals => N^d subcubes.

Take one point in each of the subcubes so that being projected to lower dimensions points do not overlap

Latin Hypercube sampling

$\{\pi_k\},\ k = 1,\dots,n$ - independent random permutations of $\{1,\dots,N\}$ each uniformly distributed over all N! possible permutations

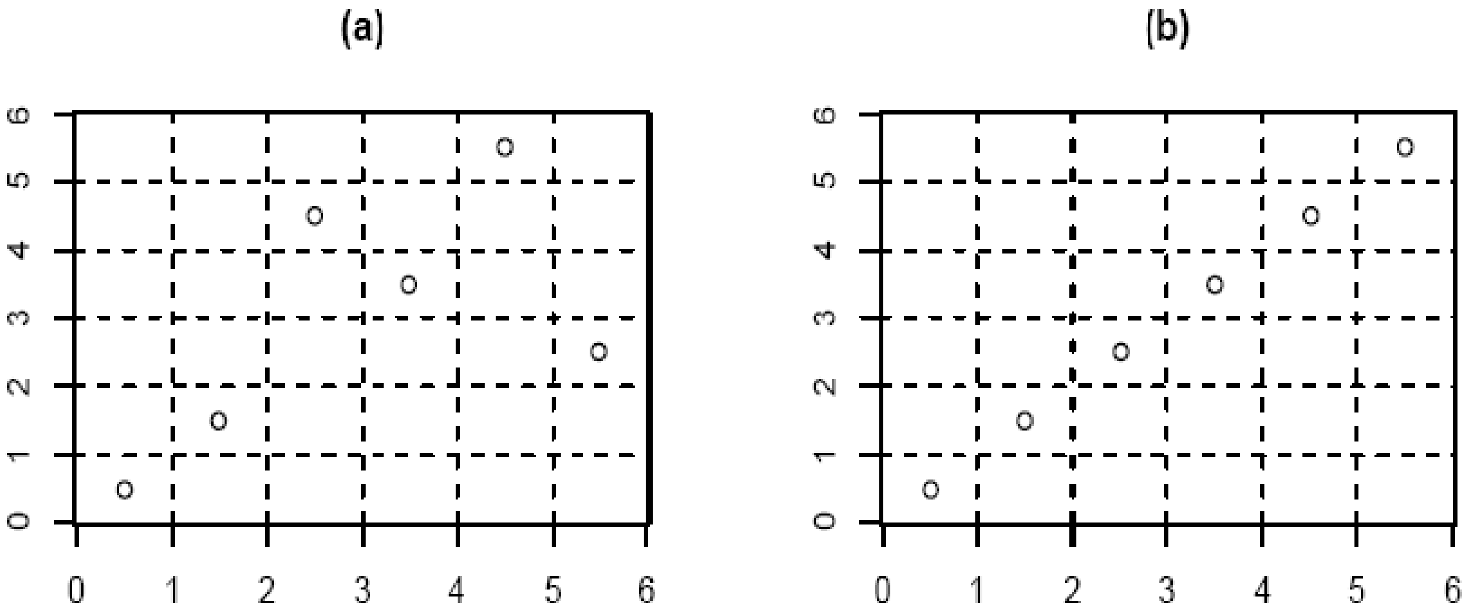
LHS coordinates: $$x_i^k = \frac{\pi_k(i) - 1 + U_i^k}{N},\ i = 1,\dots,N,\ k = 1,\dots,n$$

$$U_i^k \sim \mathrm{U}(0,1)$$

LHS is built by superimposing well stratified one-dimensional samples.

It cannot be expected to provide good uniformity properties in a n-dimensional unit hypercube.

Deficiencies of LHS sampling



1) Space is badly explored (a)

2) Possible correlation between variables (b)

3) Points can not be sampled sequentially

=> Not suited for integration

Discrepancy. Quasi Monte Carlo.

Discrepancy is a measure of deviation from uniformity:

Defintions: $Q(\vec{y}) \in H^{\mathrm{n}}$, $Q(\vec{y}) = [0, y_1) \times [0, y_2) \times \ldots \times [0, y_n)$,

$m(Q)$ – volume of Q

$$D_N^* = \sup_{Q(\vec{y}) \in H^n} \left| \frac{N_{Q(\vec{y})}}{N} - m(Q) \right|$$

Random sequences: $D_N^* \to (\ln \; \ln \; N) / N^{1/2} \sim 1 / N^{1/2}$

$D_N^* \le c(d) \dfrac{(\ln \; N)^n}{N}$ – Low discrepancy sequences (LDS)

Convergence: $\varepsilon_{QMC} = \left| I[f] - I_N[f] \right| \le V(f) D_N^*$,

$$\varepsilon_{QMC} = \frac{O(\ln \; N)^n}{N}$$

Assymptotically $\varepsilon_{QMC} \sim O(1/N) \to$ much higher than

$\varepsilon_{MC} \sim O(1/\sqrt{N})$

QMC. Sobol' sequences

$$\text{Convergence:} \quad \varepsilon = \frac{O(\ln N)^n}{N} - \text{for all LDS}$$

$$\text{For Sobol' LDS:} \ \varepsilon = \frac{O(\ln N)^{n-1}}{N}, \text{if } N = 2^k, k - \text{integer}$$

Sobol' LDS:

1. Best uniformity of distribution as N goes to infinity.
2. Good distribution for fairly small initial sets.
3. A very fast computational algorithm.

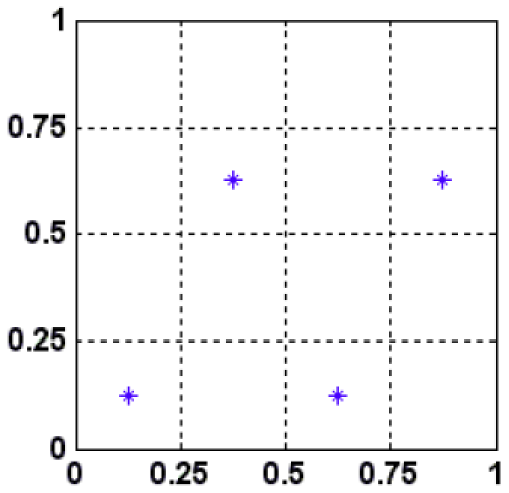
"Preponderance of the experimental evidence amassed to date points to Sobol' sequences as the most effective quasi-Monte Carlo method for application in financial engineering."

Paul Glasserman, Monte Carlo Methods in Financial Engineering, Springer, 2003

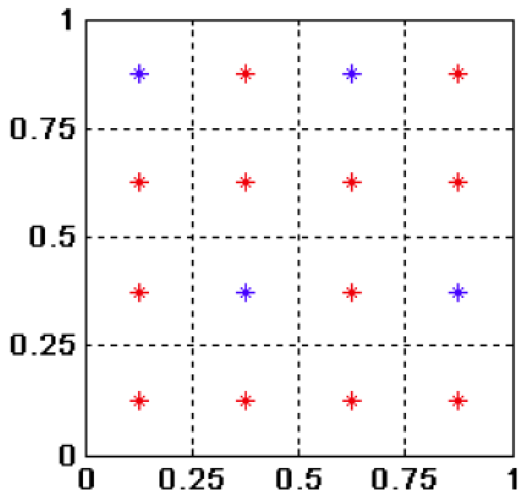
Sobol LDS. Property A and Property A’

A low-discrepancy sequence is said to satisfy Property A if for any binary segment (not an arbitrary subset) of the n-dimensional sequence of length 2^n there is exactly one point in each 2^n hyper-octant that results from subdividing the unit hypercube along each of its length extensions into half.

A low-discrepancy sequence is said to satisfy Property A’ if for any binary segment (not an arbitrary subset) of the n-dimensional sequence of length 4^n there is exactly one point in each 4^n hyper-octant that results from subdividing the unit hypercube along each of its length extensions into four equal parts.

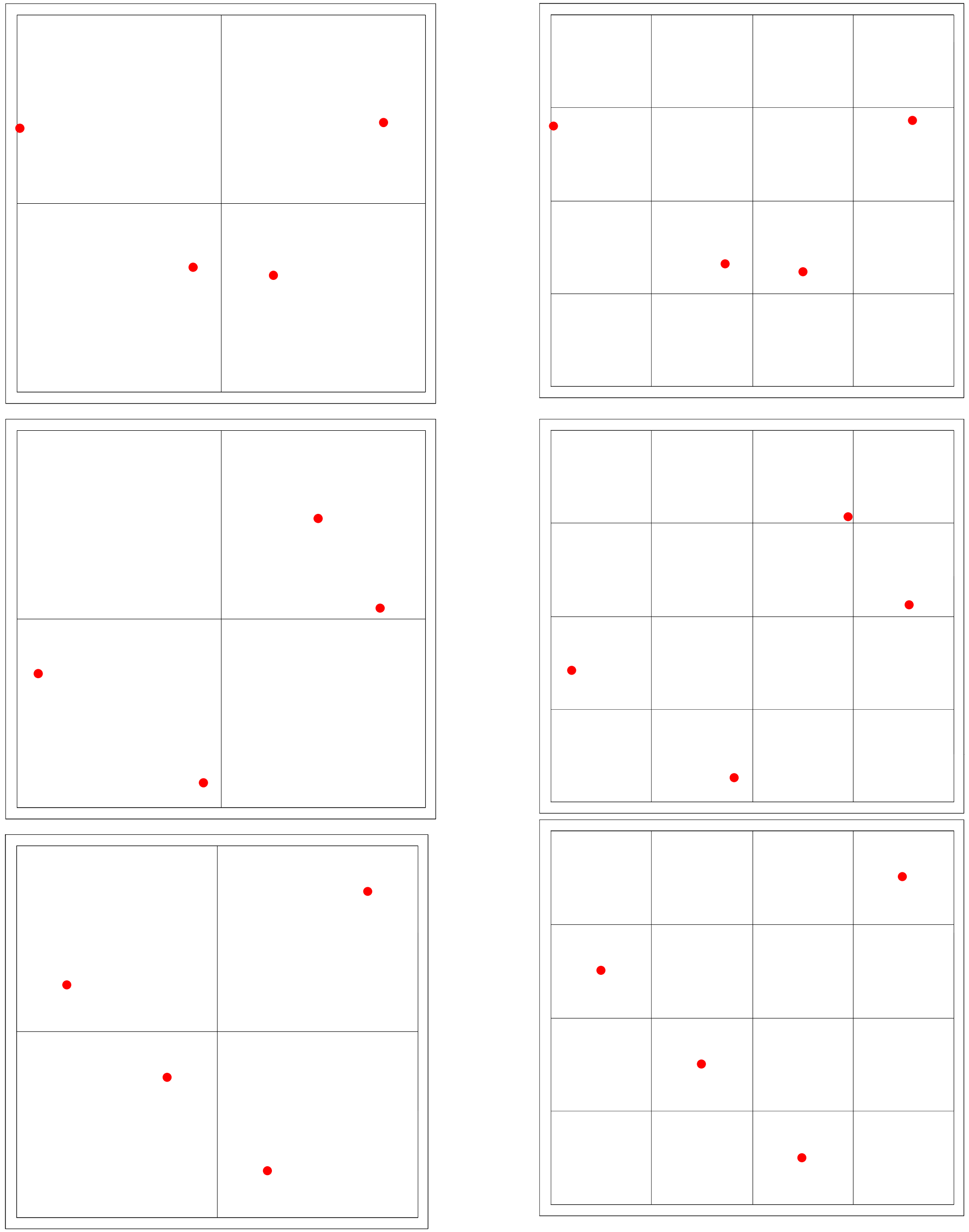


Property A



Property A’

Distributions of 4 points in two dimensions

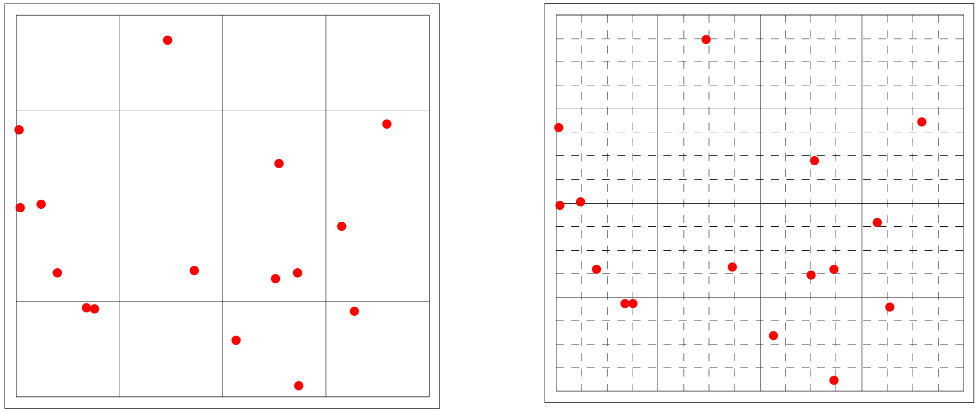


	Property A
MC ->	No
LHS ->	No
Sobol' ->	Yes

Distributions of 16 points in two dimensions

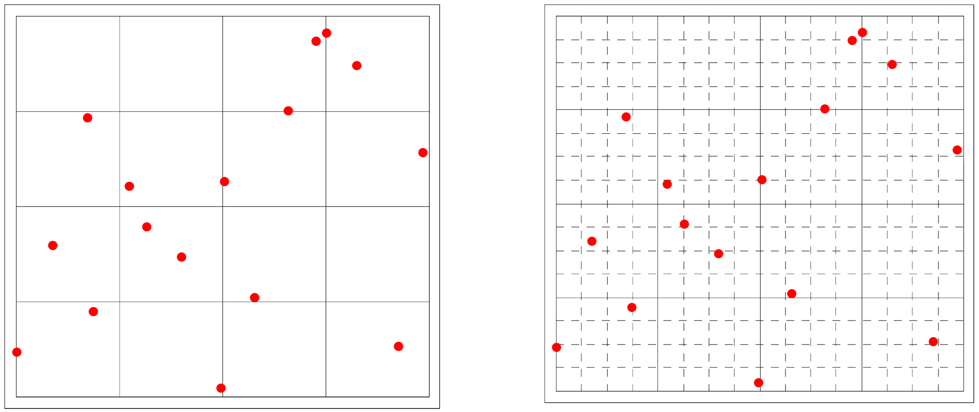
Property A'

MC ->



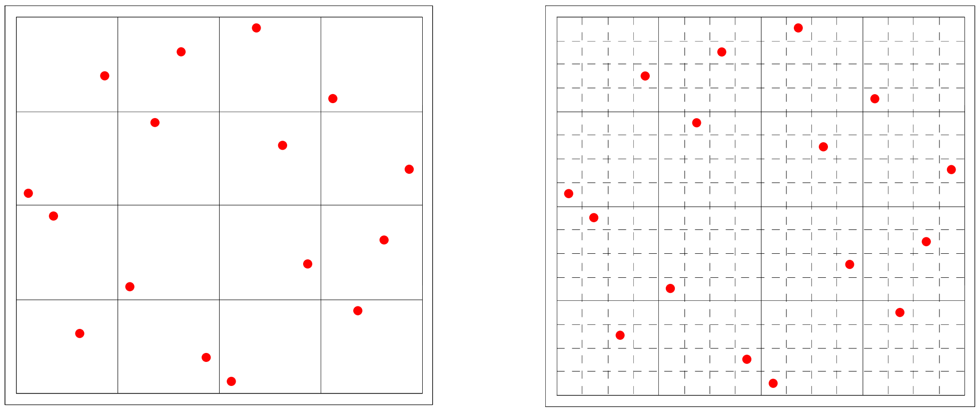
No

LHS ->



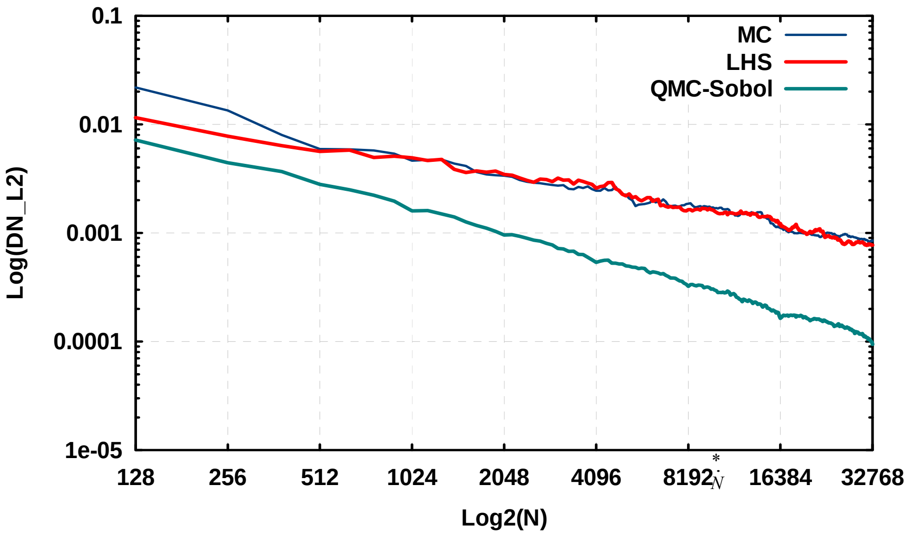
No

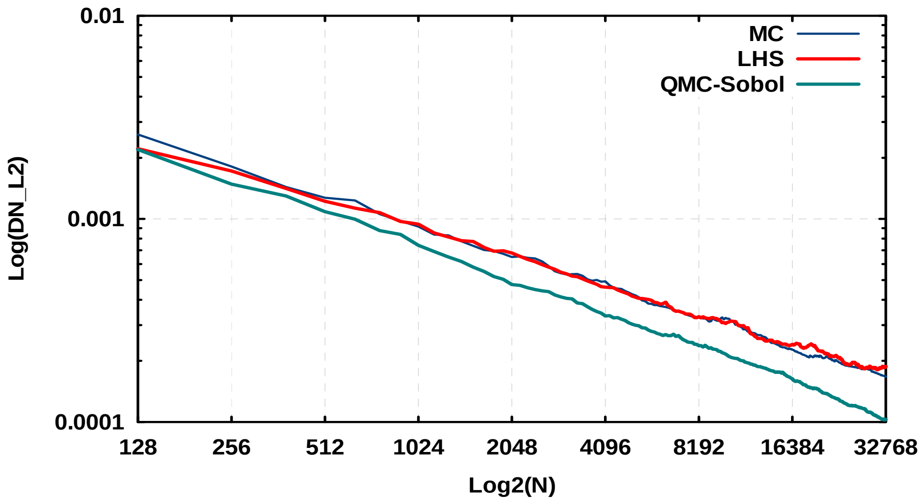
Sobol' ->



Yes

Comparison of Discrepancy I.
Low Dimensions





Use standard MC and ,

LHS generators

Sobol' sequence generator:

SobolSeq:

Sobol' sequences satisfy

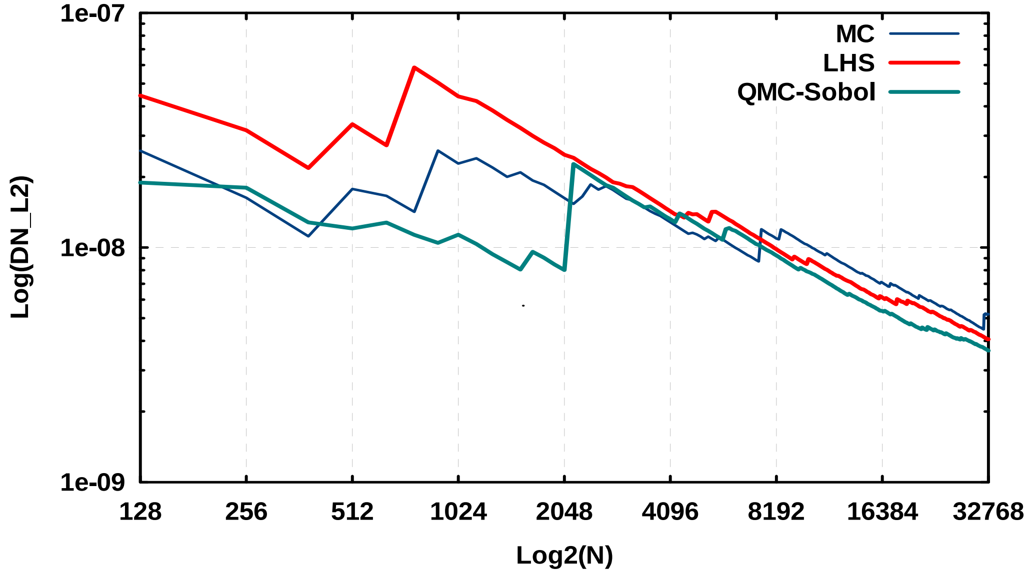
Properties A and A'

www.broda.co.uk

Result:

QMC in low dimensions shows much smaller discrepancy than MC and LHS

Comparison of Discrepancy II
High Dimensions



All sampling methods in high-dimensions have comparable discrepancy

Do QMC methods loose their efficiency in higher dimensions ?

$$\varepsilon_{QMC} = \frac{O(\ln N)^{n}}{N}$$

Assymptotically $\varepsilon_{QMC} \sim O(1/N)$

but ε_{QMC} increseas with N until $N^{*} \approx \exp(n)$

$n = 50, N \approx 5\ 10^{21}$ – not achievable for practical applications

Is QMC better than MC and LHS in higher dimensions ($\geq$ 20)

ANOVA decomposition and Sensitivity Indices

Consider a model

x is a vector of input variables

f(x) is integrable

$$Y = f(x)$$
$$x = (x_1, x_2, \dots, x_k)$$
$$0 \le x_i \le 1$$

ANOVA decomposition:

$$Y = f(x) = f_0 + \sum_{i=1}^{k} f_i(x_i) + \sum_i \sum_{j>i} f_{ij}(x_i, x_j) + \dots + f_{1,2,\dots,k}(x_1, x_2, \dots, x_k),$$
$$\int_0^1 f_{i_1 \cdots i_s}(x_{i_1}, , \dots, x_{i_s}) \quad dx_{i_k} = 0, \ \ \forall k, \ 1 \le k \le s$$

Variance decomposition:

$$\sigma^2 = \sum_i \sigma_i^2 + \sum_{i,j} \sigma_{ij}^2 + \dots \sigma_{1,2,\dots,n}^2$$

Sobol' SI:

$$1 = \sum_{i=1}^{k} S_i + \sum_{i<j} S_{ij} + \sum_{i<j<l} S_{ijl} + \dots + S_{1,2,\dots,k}$$

Sobol' Sensitivity Indices (SI)

- **Definition:** $$\boxed{S_{i_1\ldots i_s} = \sigma^2_{i_1\ldots i_s} / \sigma^2}$$

$$\sigma^2_{i_1\ldots i_s} = \int_0^1 f^2_{i_1\ldots i_s}\left(x_{i_1},\ldots,x_{is}\right)dx_{i_1},\ldots,x_{is}$$ ***- partial variances***

$$\sigma^2 = \int_0^1 \left(f\left(x\right) - f_0\right)^2 dx$$ ***- total variance***

- ***Sensitivity indices for subsets of variables:*** $x = \left(y, z\right)$

$$\sigma_y^2 = \sum_{s=1}^{m} \sum_{\left(i_1\langle\ldots\langle i_s\right)\in K} \sigma^2_{i_1,\ldots,i_s}$$

Total variance for a subset: $$\left(\sigma_y^{tot}\right)^2 = \sigma^2 - \sigma_z^2$$

Corresponding global sensitivity indices:

$$S_y = \sigma_y^2 / \sigma^2, \qquad S_y^{tot} = \left(\sigma_y^{tot}\right)^2 / \sigma^2.$$

Effective dimensions

Let $|u|$ be a cardinality of a set of variables u.

The effective dimension of $f(x)$ in the superposition sense is the smallest integer d_S such that

$$\sum_{0<|u|<d_S} S_u \geq (1-\varepsilon), \ \varepsilon << 1$$

It means that $f(x)$ is almost a sum of d_S-dimensional functions.

The function $f(x)$ has effective dimension in the truncation sense d_T if

$$\sum_{u \subseteq \{1,2,\ldots,d_T\}} S_u \geq (1-\varepsilon), \ \varepsilon << 1$$

Important property: $d_S \leq d_T$

Example: $f(x) = \sum_{i=1}^{n} x_i \rightarrow d_S = 1, \ d_T = n$

Classification of functions

Type A. Variables are not equally important

$$\frac{S_y^T}{n_y} >> \frac{S_z^T}{n_z} \leftrightarrow d_T << n$$

Type B,C. Variables are equally important

$$S_i \approx S_j \leftrightarrow d_T \approx n$$

Type B. Dominant low order indices

$$\sum_{i=1}^{n} S_i \approx 1 \leftrightarrow d_S << n$$

Type C. Dominant higher order indices

$$\sum_{i=1}^{n} S_i << 1 \leftrightarrow d_S \approx n$$

When LHS is more effective than MC ?

ANOVA: $f(x) = f_0 + \sum_i f_i(x_i) + r(x)$

$r(x)$ – high order interactions terms

LHS: $E(\varepsilon_{LHS}^2) = \frac{1}{N}\int_{H^n}[r(x)]^2 dx + O(\frac{1}{N})$ (Stein, 1987)

MC: $E(\varepsilon_{MC}^2) = \frac{1}{N}\sum_i \int_{H^n}[f_i(x_i)]^2 dx + \frac{1}{N}\int_{H^n}[r(x)]^2 dx + O(\frac{1}{N})$

if $\int_{H^n}[r(x)]^2 dx$ is small $\Leftrightarrow d_S$ (Type B functions)

$\rightarrow \quad E(\varepsilon_{LHS}^2) < E(\varepsilon_{MC}^2)$

Classification of functions

Function type	Description	Relationship between S_i and S_i^{tot}	d_T	d_S	QMC is more efficient than MC	LHS is more efficient than MC
A	A few dominant variables	$S_y^{tot}/n_y >> S_z^{to} / n_z$	$<< n$	$<< n$	Yes	No
B	No unimportant subsets; only low-order interaction terms are present	$S_i \approx S_j, \forall\ i, j$ $S_i / S_i^{tot} \approx 1, \forall$ i	$\approx n$	$<< n$	Yes	Yes
C	No unimportant subsets; high-order interaction terms are present	$S_i \approx S_j, \forall\ i, j$ $S_i / S_i^{tot} << 1, \forall\ i$	$\approx n$	$\approx n$	No	No

How to monitor convergence of MC, LHS and QMC calculations ?

The root mean square error is defined as

$$\varepsilon = \left(\frac{1}{K} \sum_{k=1}^{K} (I_d - I_N^k)^2 \right)^{1/2}$$

K is a number of independent runs

MC and LHS: all runs should be statistically independent (use a different seed point).

QMC: for each run a different part of the Sobol' LDS was used (start from a different index number).

The root mean square error is approximated by the formula

$$cN^{-\alpha},\ 0 < \alpha < 1$$

$$\text{MC:}\quad \alpha \approx 0.5$$

$$\text{QMC:}\ \alpha \leq 1$$

$$\text{LHS:}\quad \alpha \sim ?$$

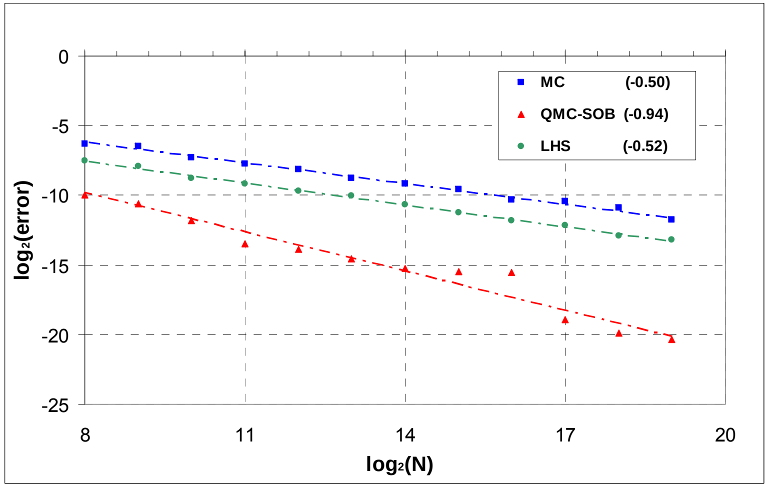
Integration error vs. N. Type A

(a) $f(x) = \sum^{n}_{j=1}(-1)^{i}\prod^{i}_{j=1} x_j$, n = 360, (b) $f(x) = \prod^{s}_{i=1} |4x_i-2|/(1+a_i)$, n = 100

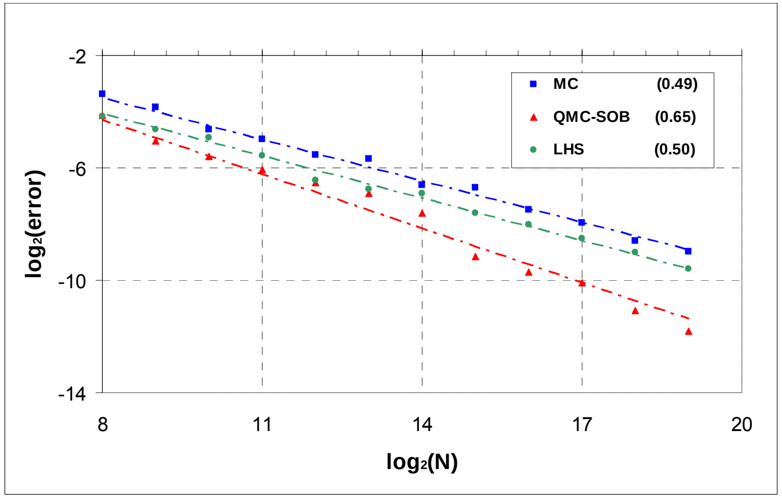
$$\varepsilon = \left(\frac{1}{K}\sum_{k=1}^{K}(I - I_N^k)^2\right)^{1/2}$$

$$\varepsilon \sim N^{-\alpha},\ 0<\alpha<1$$

(a)



(b)



$$\frac{S_y^T}{n_y} >> \frac{S_z^T}{n_z} \leftrightarrow d_T << n$$

Integration error. Type A

$$\varepsilon = \left(\frac{1}{K} \sum_{k=1}^{K} (I - I_N^k)^2 \right)^{1/2}$$

$$\varepsilon \sim N^{-\alpha}, \; 0 < \alpha < 1$$

$$\frac{S_y^T}{n_y} >> \frac{S_z^T}{n_z} \leftrightarrow d_T << n$$

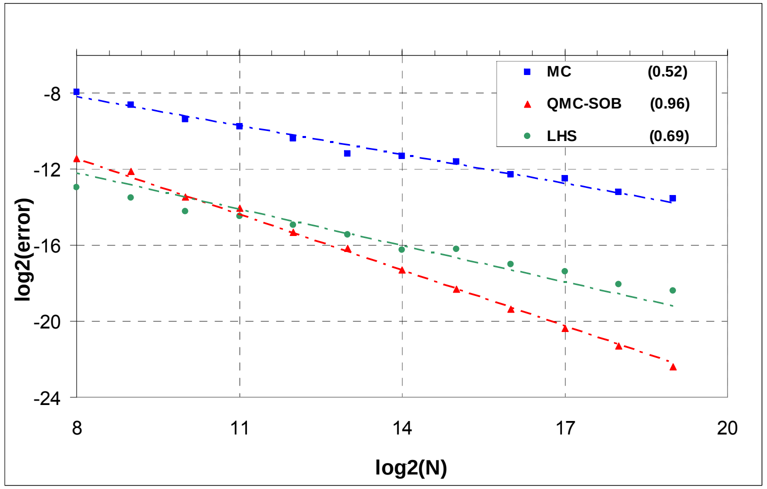
Index	Function	Dim n	Slope MC	Slope QMC	Slope LHS
1A	$\sum_{i=1}^{n} (-1)^i \prod_{j=1}^{i} x_j$	360	0.50	0.94	0.52
2A	$\prod_{i=1}^{n} \frac{\lvert 4x_i - 2 \rvert + a_i}{1 + a_i}$ $a_1 = a_2 = 0$ $a_3 = \ldots = a_{100} = 6.52$	100	0.49	0.65	0.50

Integration error vs. N. Type B

Dominant low order indices

$$\sum_{i=1}^{n} S_i \approx 1 \leftrightarrow d_S << n$$

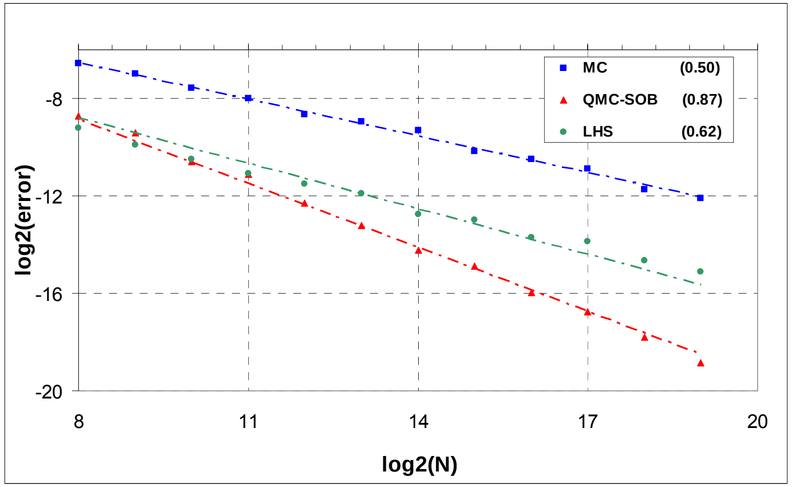
(a)



$$f(x) = \prod_{i=1}^{n} \frac{n - x_i}{n - 0.5}$$

$$n = 360$$

(b)



$$f(x) = \prod_{i=1}^{n} (1 + 1/n) x_i^{1/n}$$

$$n = 360$$

Integration error. Type B functions

Dominant low order indices

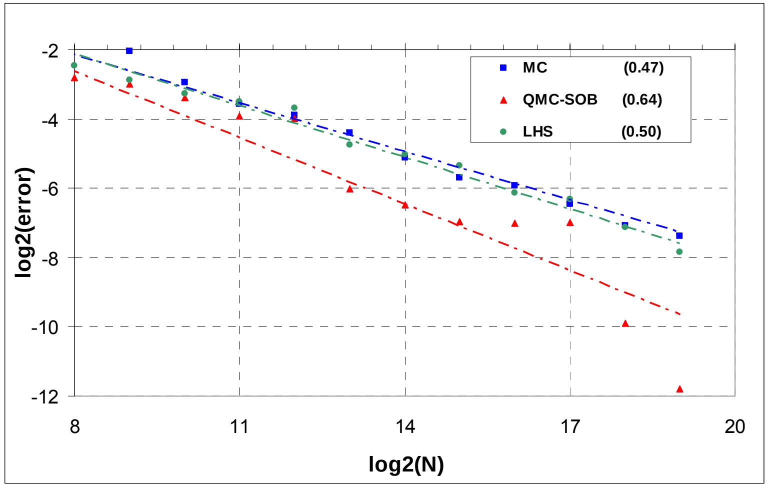
$$\sum_{i=1}^{n} S_i \approx 1 \leftrightarrow d_S << n$$

Index	Function	Dim n	Slope MC	Slope QMC	Slope LHS
1B	$\prod_{i=1}^{n} \frac{n - x_i}{n - 0.5}$	30	0.52	0.96	0.69
2B	$\left(1 + \frac{1}{n}\right)^n \prod_{i=1}^{n} \sqrt[n]{x_i}$	30	0.50	0.87	0.62
3B	$\prod_{i=1}^{n} \frac{\lvert 4x_i - 2 \rvert + a_i}{1 + a_i}$ $a_i = 6.52$	30	0.51	0.85	0.55

The integration error vs. N. Type C

Dominant higher order indices: $\sum_{i=1}^{n} S_i << 1 \leftrightarrow d_S \approx n$

(a)

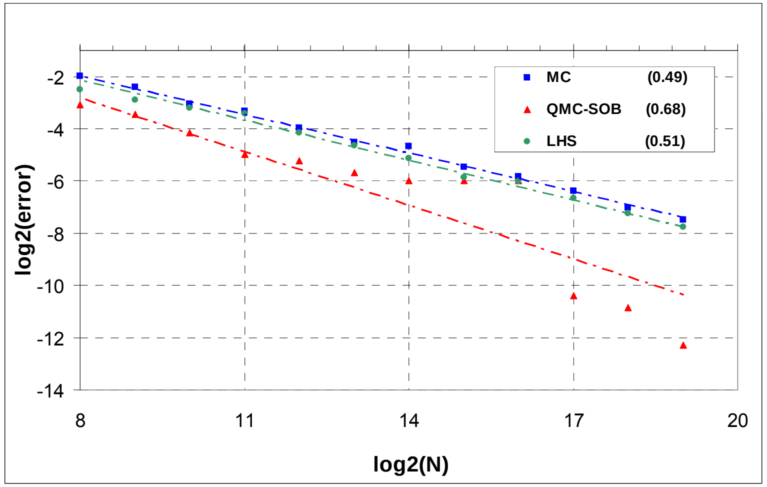


$$f(x) = \prod_{i=1}^{n} \frac{|4x_i - 2| + a_i}{1 + a_i}, a_i = 0$$

$$\rightarrow \prod_{i=1}^{n} |4x_i - 2|$$

$$n = 10$$

(b)



$$f(x) = (1/2)^{1/n} \prod_{i=1}^{n} x_i$$

$$n = 10$$

Integration error for type C functions

Dominant higher order indices

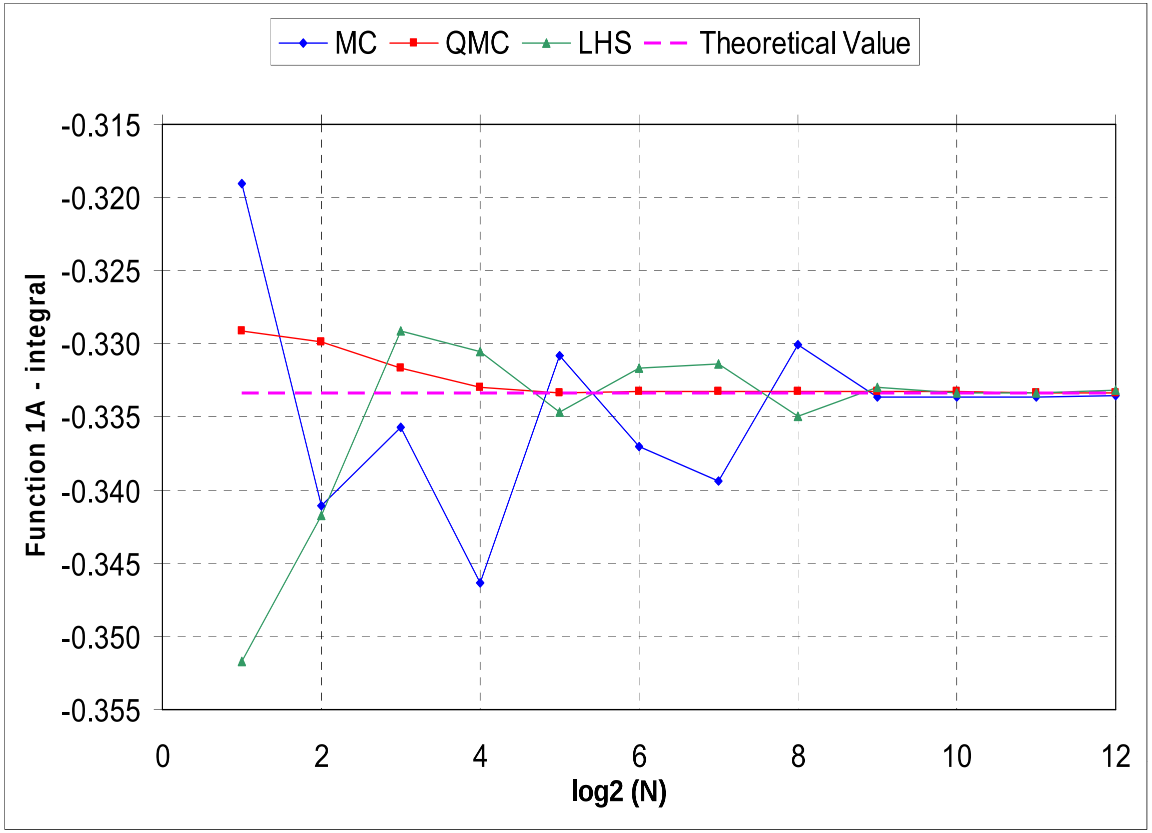
$$\sum_{i=1}^{n} S_i << 1 \leftrightarrow d_S \approx n$$

Index	Function	Dim n	Slope MC	Slope QMC	Slope LHS
1C	$\prod_{i=1}^{n} \lvert 4x_i - 2 \rvert$	10	0.47	0.64	0.50
2C	$(2)^n \prod_{i=1}^{n} x_i$	10	0.49	0.68	0.51

The integration error vs. N. Function 1A

$$\sum_{i=1}^{n}(-1)^{i}\prod_{j=1}^{i}x_j,$$

$$n = 360$$



QMC: convergence is monotonic
MC and LHS: convergence curves are oscillating

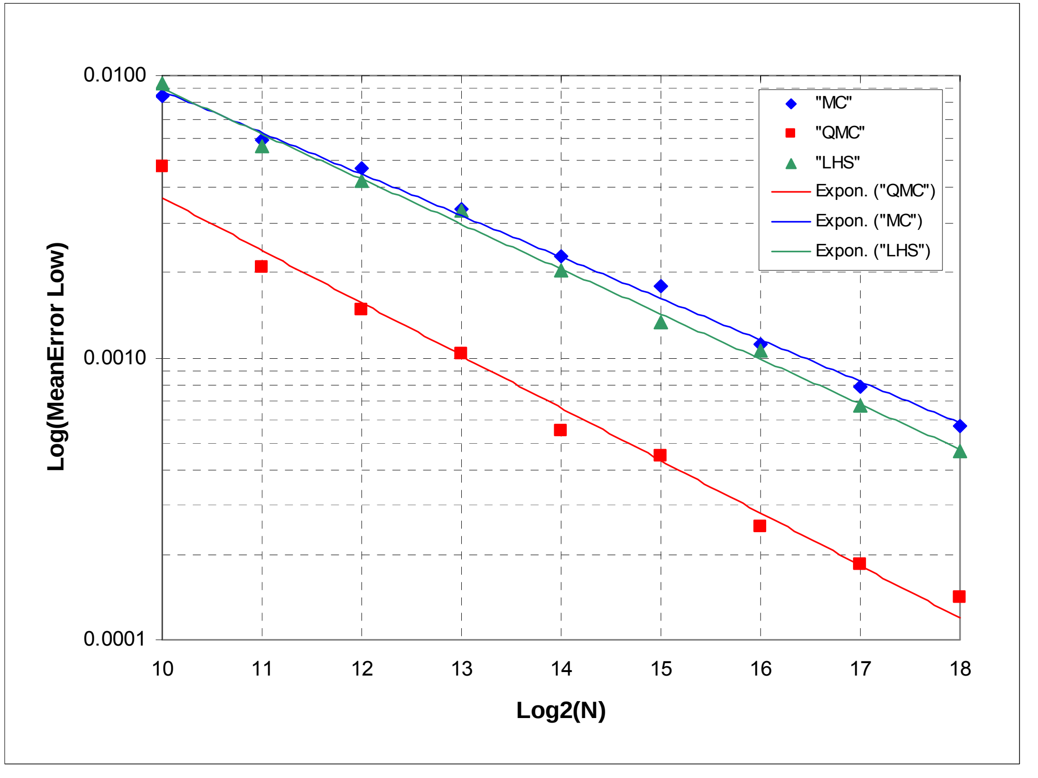
QMC is 30 times faster than MC and LHS

LHS: it is not possible to incrementally add a new point while keeping the old LHS design

Evaluation of quantiles I. Low quantile

Distribution $f(x) = \sum_{i=1}^{n} x_i^2,$ dimension n = 5.

$x_i \sim N(0,1)$ are independent standard normal variates

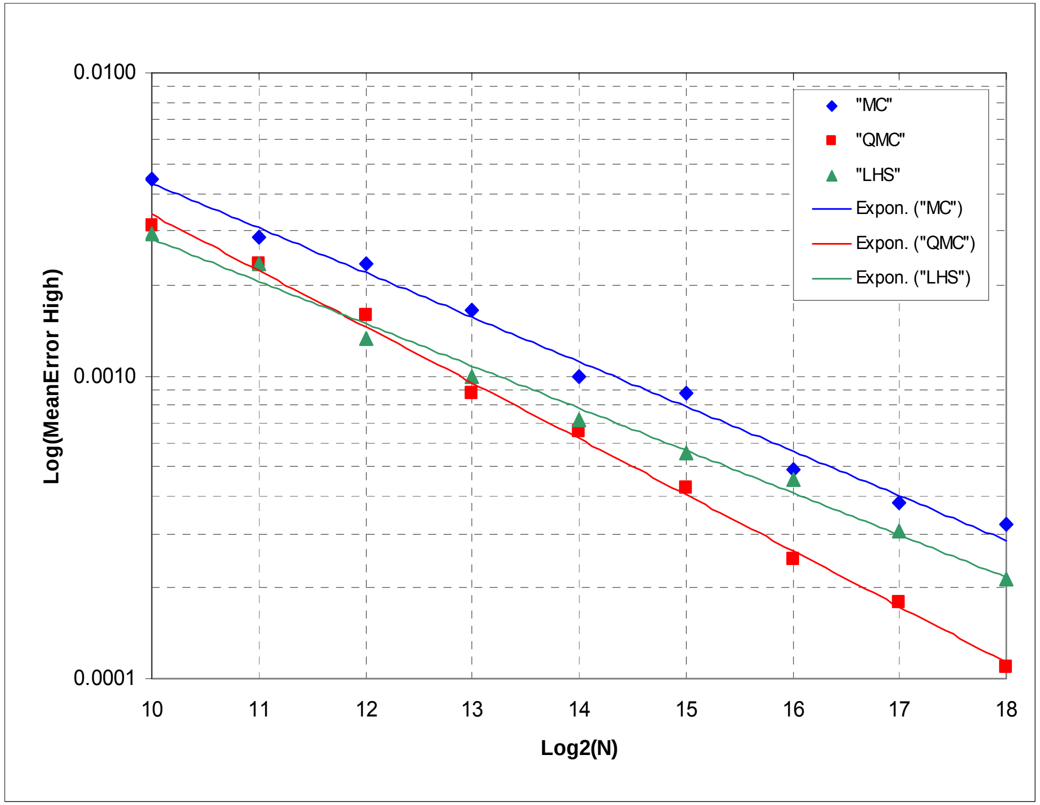


Low quantile (percentile for the cumulative distribution function) = 0.05
A superior convergence of the QMC method

Evaluation of quantiles II. High quantile

Distribution $f(x) = \sum_{i=1}^{n} x_i^2,$ dimension n = 5.

$x_i \sim N(0,1)$ are independent standard normal variates



High quantile (percentile for the cumulative distribution function) = 0.95

QMC convergences faster than MC and LHS

Summary

Sobol' sequences possess additional uniformity properties which MC and LHS techniques do not have (Properties A and A').

Comparison of L_2 discrepancies shows that the QMC method has the lowest discrepancy in low dimensions (up to 20).

QMC method outperforms MC and LHS for types A and B functions (problems with low effective dimensions)

LHS method outperforms MC only for type B functions.

QMC remains the most efficient method among the three techniques for non-uniform distributions